Open Access Journal of Clinical Trials

Dovepress
open access to scientific and medical research

Open Access Full Text Article

PERSPECTIVES

Clinical data management: Current status, challenges, and future directions from industry perspectives

This article was published in the following Dove Press journal:
Open Access Journal of Clinical Trials
19 June 2010
Number of times this article has been viewed

Zhengwu Lu[1]
Jing Su[2]

[1]Smith Hanley Consulting, Houston, Texas; [2]Department of Chemical Engineering, University of Massachusetts, Amherst, MA, USA

Abstract: To maintain a competitive position, the biopharmaceutical industry has been facing the challenge of increasing productivity both internally and externally. As the product of the clinical development process, clinical data are recognized to be the key corporate asset and provide critical evidence of a medicine's efficacy and safety and of its potential economic value to the market. It is also well recognized that using effective technology-enabled methods to manage clinical data can enhance the speed with which the drug is developed and commercialized, hence enhancing the competitive advantage. The effective use of data-capture tools may ensure that high-quality data are available for early review and rapid decision-making. A well-designed, protocol-driven, standardized, site workflow-oriented and documented database, populated via efficient data feed mechanisms, will ensure regulatory and commercial questions receive rapid responses. When information from a sponsor's clinical database or data warehouse develops into corporate knowledge, the value of the medicine can be realized. Moreover, regulators, payer groups, patients, activist groups, patient advocacy groups, and employers are becoming more educated consumers of medicine, requiring monetary value and quality, and seeking out up-to-date medical information supplied by biopharmaceutical companies. All these developments in the current biopharmaceutical arena demand that clinical data management (CDM) is at the forefront, leading change, influencing direction, and providing objective evidence. Sustaining an integrated database or data repository for initial product registration and subsequent postmarketing uses is a long-term process to maximize return on investment for organizations. CDM should be the owner of driving clinical data-cleaning process in consultation with other stakeholders, such as clinical operations, safety, quality assurance, and sites, and responsible for building a knowledge base to add potential value in assisting further study designs or clinical programs. CDM needs to draw on a broad range of skills such as technical, scientific, project management, information technology (IT), systems engineering, and interpersonal skills to tackle, drive, and provide valued service in managing data within the anticipated e-clinical age. Commitment to regulatory compliance is required in this regulated industry; however, a can-do attitude with strong willingness to change and to seek ways to improve CDM functions and processes proactively are essential to continued success and to ensure quality data-driven productivity.

Keywords: clinical trials, data management, standard, efficacy, safety, clinical systems, clinical data, electronic data-capturing

Correspondence: Zhengwu Lu
1111 Weyburn LN#29, San Jose, CA 95129, USA
Tel +1 201 233 0738
Fax +1 201 949 4085
Email zhengwu.lu@ieee.org

Introduction

It is recognized that clinical data are key corporate assets in today's biopharmaceutical industry, and that turning data into meaningful information is a critical core function for sponsor firms to make faster and more flexible assessments of compounds in

development, design better clinical protocols when tailoring the appropriate target population with a specific indication, and enable innovative study initiatives and new clinical programs to ensure a robust clinical product pipeline. Clinical data management (CDM) is a vital cross-functional vehicle in clinical trials to ensure high-quality data are captured by sites staff through paper case report form (CRF) or electronic case report form (eCRF) and available for early review. The integrity and quality of data being collected and transferred from study subjects to a clinical data management system (CDMS) must be monitored, maintained, and quantified to ensure a reliable and effective base for not only new drug application (NDA) submission and clinical science reports but also corporate clinical planning, decision-making, process improvement, and operational optimization.

The gradually increasing use of electronic data-capturing (EDC) technology and eCRF to collect data in clinical trials has grown in recent years and has affected the activities of clinical research operations for industry sponsors, contract research organizations (CROs), and clinical sites.[1–3] EDC technology must comply with applicable regulatory requirements and offer flexible, configurable, scalable, and auditable system features.[4] Transitioning from paper-based data collection (PDC) to EDC systems has produced many benefits, ie, easing the burden associated with organizing paper CRF work and greatly reducing the time, cost, and stress required in bringing a product to market through technology-enabled efficiency improvement, such as the quick and robust interactive voice response system (IVRS) supported and integrated auto casebook creation, early data availability, and fast database lock via Internet-based user interface. Although EDC technologies offer advantages over traditional paper-based systems, collecting, monitoring, coding, reconciling, and analyzing clinical data. often from multiple sources, can be challenging.

To realize the full potential of technology advantage in clinical research, both sponsor and site users need to change the way their offices and days are organized, how they enter and retrieve patient information, the process by which they issue, answer, or close queries, the standard operating procedures (SOPs), work practices, guidelines, and business documents, and the ways in which they relate to colleagues and CROs and interact with their patients.[2] To address the challenges of the e-clinical environment and reap the benefits of technology, business re-engineering, organizational realignment, and management commitment are required to ensure that biopharmaceutical firms adapt to a culture embracing technology, and develop or revise existing legacy procedures to accommodate the re-engineered e-clinical processes and procedures.[5]

EDC technology will not guarantee the quality and integrity of collected data. The main source of error in PDC trials was when data were extracted from patient medical records and transcribed to the CRF. This activity stays the same with EDC, where data are extracted from the same source, entered into eCRF and then saved into the CDMS. To enable high integrity and quality data for analysis and submission using EDC, data managers and all related functional members, including CROs, must understand how this new technology, related clinical systems, and processes affect data quality.

Consequently, biopharmaceutical companies have been undergoing major changes in reassessing their IVRS, CDMS, clinical trial management system (CTMS), and clinical safety system (CSS) to accommodate the growing needs and demands. Multiple vendors supply various such systems in commercial software packages. Challenges and improvement opportunities exist in customization, configuration, or integration among the adopted systems for a sponsor e-clinical environment to engender clinical efficiencies and quality improvement. This presents exciting times in which sponsors can connect themselves to clinical sites more dynamically to drive clinical operation and site productivity with e-clinical solutions, such as clinical web portals. To maximize return using technologies, sponsor firms need to evaluate and carefully select technology vendors, platforms, or applications to address the unique requirements of clinical trials-investigator gathered data, patient-entered e-diary data, adverse event reporting, and text reminders for patients. With incorporated clinical data standards such as the Clinical Data Interchange Standards Consortium (CDISC), these interconnected systems will present the future vision of integrated data and systems, and produce much enhanced value to the corporation. Further, achieving effective interoperability between electronic health care records (eHR) and CDMS is highly desirable for many parties, yet a number of legal, technical, and ethical barriers mean that this connectivity remains largely a vision at present. In this technical viewpoint, the authors seek to clarify some of the issues that are central to current discussions about CDM, focusing on topics critical to biopharmaceutical companies having compounds in clinical development for human use.

This paper is prepared from industry perspectives to present and analyze the cross-functional role of CDM, current status of PDC and EDC, benefits of new processes and technologies, challenges, and risks associated with EDC, based on systematic overview. This article addresses four questions: What are CDM and the role of data managers? What do we do in the coexisting world of PDC and EDC? What challenges are out there preventing the widespread

usage of EDC technology? What does the future hold for CDM in conducting EDC studies?

What is clinical data management?

A biopharmaceutical industry definition

CDM is defined as "the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets"[6] in the clinical trial arena. With its diverse connectivity, cross-functional features, and a wide range of responsibilities, CDM has come a long way in the past two decades, and is a recognized profession with increasingly realized importance within and outside biopharmaceutical research and development. As complex and dynamic as the profession is, CDM globally continues to grow into a firmly established discipline in its own right, focuses on managing clinical trial-related data as a valuable resource, and is becoming a career that requires multiple skill sets, such as a background of sound clinical skills, scientific rigor, information technology, systems engineering, and strong communications ability. With the continued global harmonization of clinical research and introduction of regulatory-mandated electronic submission in the industry, it is critical to understand, appreciate, work within the framework of global clinical development, and apply standards in the development and execution of architectures, policies, practices, guidelines, and procedures that properly manage the full clinical data lifecycle needs of an enterprise. This definition is fairly broad and encompasses a number of professions which may not have direct technical contact with lower-level aspects of data management, such as relational database management. Many other topics, processes, and procedures are also relevant, including:

- Data governance, such as standards management, SOPs, and guidelines
- Data architecture, analysis, and design including data modeling for potential clinical data repository or warehouse
- Database management including data maintenance, administration, and data mapping across related clinical or external systems
- Data security management including data access, archiving, privacy, and security
- Data quality management including query management, data integrity, data quality, and quality assurance
- Reference and master data management including data integration, external data transfer, master data management, reference data
- Data warehousing and business intelligence (BI) management including tools, data mining, and ETL (extract, transform, and load)
- Document, record, and content management
- Metadata management, ie, metadata definition, discovery, publishing, metrics, and standardization.[7]

Clinical data management perspectives

CDM has evolved and will continue to develop in response to the special cross-functional needs and according to the particular strengths of e-clinical research advances due to much enhanced clinical harmonization, global standardization, and expected clinical systems interoperability initiatives. The future is not what it used to be, and will undergo many anticipated reality checks. CDM professionals once optimistically predicted that EDC technology would radically increase efficiency by reducing the amount of paper documentation associated with clinical trials, and streamline the CDM process considerably. Indeed, some sponsor companies have realized some claimed clinical efficiencies with planned long-term cost savings, but not all of them do so well. It is not uncommon to see sponsor companies spending a large resource and investment to establish an electronic documentation system, such as Electronic Documentum, to store study-related documents while still maintaining a concurrent manual paper filing system. It seems a reasonable reality that the current clinical studies are operated in both traditional PDC-based and EDC-supported environments by sponsors and/or CROs with differential levels of automation. The speed at which paper mountains accumulate may have been reduced by some sponsor companies; however, adoption of an electronic document management or clinical trial management system seems unable to eliminate the document piling. Therefore, successful implementation and integration of EDC technology with other key clinical systems depends as much on managing change as it does on clinical science and technology itself, and changes, especially organizational ones, have never been easy for sponsor e-clinical solutions implementation.[2] To realize the full potential of EDC technology in e-clinical research, both sponsor and site personnel need to make logistic reorganizational changes in their offices and surroundings, in entering and retrieving clinical information, in managing the issuance or closure of queries, in interacting and dealing with other stakeholders such as colleagues, CROs, and study subjects, and, most importantly, in gaining an understanding of the technology advantages and limits to achievement of business objectives.

Electronic solutions in clinical data management

Technology-driven strategies and initiatives have the potential to alleviate the significant pressure to market a medicine as early in the patent life as possible to maximize the period without competition, both to increase total revenue and to shorten the time to market sales. The increase in regulatory requirements and competition seen in the recent years, coupled with reforms in health care services, has presented extreme challenges for the biopharmaceutical industry, suggesting the need for sponsor companies to invest significantly in technological solutions and add an additional emphasis on business process re-engineering and improvement to engender long-term clinical efficiencies and cost benefits. In this environment, the effectiveness of the clinical data management function is crucial to substantiate early approval for a new product launch and subsequent successful marketing. Delay, deficiency, or quality issues in the CDM process can be costly. Further, speed is not enough by itself and success needs to be achieved with other quality attributes. There is an ever-increasing demand for sponsors, including CROs, to strike the right balance between time, cost, process, and quality in conducting all clinical studies. Applying e-clinical technology, including EDC, in such a context is the anticipated industry trend and will continue to offer superior benefits to sponsors as collaboration, standardization initiatives, and technology innovation are constantly geared towards more and wider technology adoption.

Status of data management in clinical studies

Slow yet increasing EDC adoption combined with EDC technology improvement has demonstrated the reality and complexity of implementing re-engineered e-clinical processes along with new technology introduction. There is still the presence of PDC in a large number of sponsor firms, especially in Phase I clinical studies or studies sponsored by small-sized or start-up firms. Medium or large biopharmaceutical firms are tending to move into EDC, or have accumulated implementation expertise with the technology and associated e-clinical systems. It is not surprising that the traditional PDC and evolving EDC may coexist for a sponsor or CRO. To address the clinical operational needs, a sponsor firm or CRO may have a different set of procedures, standard work practices, guidelines, or business documents for PDC and EDC. Some sponsors may outsource the PDC data management functions to CROs in a complete fashion. Other sponsors may take a combinational approach whereby they would have an internal core team design the CRFs and come up with varied edit check specifications, but seek CROs to build the database and program those checks. To ensure that a standardized set of forms and edit checks are applied for cross-therapeutic clinical studies, sponsor firms must have the proper oversight and expertise to drive CRO data management or database design deliverables. There also seems to be an evolving trend whereby sponsor firms separate clinical database design (CRF or eCRF) and deployment functions into a specific unit from the CDM group due to the increasing sophistication of technology improvement, innovation, or clinical systems integration. It is also common for a different clinical programming unit to be set up for programming edit checks, listings, or reports for different functional groups. Increasing EDC computerization has enabled a paperless environment where key study variables based on protocols and electronic querying need to be transmitted between a clinic and a sponsor via a web browser entry. An independent CDM organizational unit with data managers designated to various therapeutic areas seems to be more beneficial to sponsors in terms of standardization, systems integration, and process consolidation than multiple CDM units affiliated with different therapeutic functions.

Scope of clinical data management

It is now a known fact that the scope of data capture, CRF design, and CDM activity vary widely between different companies engaging in clinical studies. For small-size entities, traditional data entry from paper CRF at a central location or outsourced CRO may still be the most effective stategy when all factors are taken into consideration. Larger companies have turned to EDC technology to deal with ongoing clinical study challenges, and long-term benefits of pursuing EDC-enabled global strategies are being realized gradually. The associated changes in the CDM process and ensuing reorganizational structuring indicate that the roles of those employed in CDM become increasingly blurred with those of their colleagues in clinical monitoring, quality assurance, and application development.[8] Moreover, the pace of technology development or optimization may be so rapid that additional consideration is required for any company planning to invest in new hardware and software for EDC technology in a changing operational environment.

Roles and responsibilities

In this mixed PDC and EDC environment, clinical data managers and CRF designers should be involved in the earliest development of the strategies and tools for data collection. Table 1 lists potential CDM key activities prior to the planning of site initiation visit for a typical study.

Through participation with the team during the design of the study, the data manager or study designer gains the necessary understanding of the required data from the protocol and the standards expected with respect to data quality. It is important for data managers or study designers to understand the varied sources of the data and the form in which the data will be retrieved, ie, hospital records, laboratory test results, insurance and government records, private physician records, or e-diaries/patient-reported outcomes. It is increasingly recognized that the design of the CRF or eCRF is a key quality step in ensuring the data required by the protocol, regulatory compliance and/or safety needs/comments, study scientific-specific hypothesis attributes, site work flow, and cross-checking of data items within a form or across different forms are addressed. CRF design is an interdisciplinary system engineering process requiring not only technical skills in utilizing the information technology (IT) tools but also expertise and scientific reasoning in the subject therapeutic areas. The original materials for this critical design are the draft yet stable clinical protocol, the corporate therapeutic unit standard forms, and clinical data acquisition standards harmonization (CDASH) guidelines. Such systems engineering work requires cross-functional team collaboration and input. It is mission critical that all functional teams including science, safety, biostatistics, regulatory compliance, and IT are represented in form review meetings and their feedback is incorporated into the revised and finalized forms. Systems development methodology and controlled process are followed for eCRF design and development to ensure regulatory requirements are met.[5] Additionally, form design must always be tailored to the majority of end users and have their work flow taken into account. Any potential ambiguity in the CRF or eCRF must be avoided.[10] In today's clinical research, the concepts and definitions are reasonably standardized. For each study, the definition of clinical terms, data entry guidelines, and data handling conventions require intensive effort and communication among all members of the study team to assure a meaningful and persistent set of data is compiled. Such information should be incorporated into written guidelines for CRF or eCRF completion. The use of the CRFs and guidelines should be thoroughly tested and reviewed by a pilot use at least among clinical data management or verification staff. Data edits such as ranges and cross-checks should be established with the participation of CDM, monitoring personnel, and scientists. This is especially important with EDC studies because the majority of such edit checks impact how queries will be issued and resolved.

Measurement of performance

The conduct of a clinical trial involves a complex interplay between many teams, with a multitude of processes taking place in the critical path of clinical product development. In the course of the study, the CDM or quality assurance team should continually assess and verify the data collection and database for completeness and consistency. No study goes exactly as planned. The measurement of performance and productivity is pivotal to drive the successful achievement of project goals, each minor milestone forging the path to the next on the road to registration. Additionally, there seems to be a common need to develop and adopt a list of primary performance indicators so that a "complete form" means the same thing across all therapeutic areas and all clinical studies, regardless whether they are PDC- or EDC-based. A sponsor may end up considering a form complete only when all required data entry has been completed, when there are no open queries, and the form has been source-verified. However, it would be reasonable for a different sponsor to drop out the source verification require-

Table 1 A list of clinical data management preinitiation activities for clinical studies in either PDC or EDC

Activity	Description
Design study	Objectives, scope, inclusion and exclusion criteria, primary endpoint(s), secondary endpoint(s), limitations, comparators (if any), project plan
Develop data collection strategy	Patient self-report via e-diary, survey and/or medical records
Design form (CRF or eCRF)	Use sponsor-designated vendor tool(s)
Prepare edit check specification	In consultation with science, biostatistics, safety, and quality assurance
Prepare data entry guideline	In consultation with CRA and CROs if needed
Design database or integration	Adopt sponsor-designated platform
Prepare study monitoring plan	Review and offer suggestions
Design record log for tracking CRF	For CRF reconciliation
Prepare study management plan	High level plans on process, communication, data handling conventions, criteria, and performance metrics

Abbreviations: PDC, paper-based data collection; EDC, electronic data-capturing; CRA, clinical research associate; CRF, case report form; eCRF, electronic case report form; CRO, contract research organization.

ment for a form to be considered complete. Clearly, having a consistent definition, data management process, and standards maintained across all studies will enable objective analysis and performance comparisons to ensure optimized efficiency and manage achievement of key milestones along the complex critical path. Performance or metrics reports indicate a measure of fitness for purpose, can be used as an overall measure of quality and productivity or work rate, and are key tools for assessing the functions of the data management, clinical project, or CROs involved. A defined set of quantitative measures can be extracted from the CDMS, presented to assist in management of the process flow and for improvement identification, and generally falls into three distinct categories,[9] ie, status reporting (measuring productivity against resources), measurement and reporting of quality, and measurement and reporting of process cycle times.

Developing and adopting such metrics reports involves collaborative efforts across multiple stakeholders. Some sponsors may utilize such performance reports to trigger CRO or site payment. Others may refer to these for assessing the performance of CDM, clinical research associate (CRA), or quality control (QC) staff. One must realize that these indicators are tools for sponsors to ensure timely delivery of high-quality data through many cross-functional groups to satisfy both good clinical practice (GCP) requirements and the statistical analysis and reporting requirements.

Challenges in clinical data management

Although EDC technology and e-clinical systems have been implemented to enhance various aspects of the data management process, implementation has not been without difficulty nor has it been improved as rapidly as many had anticipated. The pharmaceutical, biotechnology, and medical device industry, as well as academia and the government, have all started to learn about the technology advantages; some have gained implementation expertise in adopting or configuring it as a new data management tool. EDC acceptance seems strong, and there are few instances where sponsors have gone back to PDC studies when they have had the experience of EDC. Although the goal of data management will not change, ie, assurance of clean data at the end of the study, there is no doubt that data management processes will evolve with the use of EDC and e-clinical systems.

Critical clinical form design with balancing needs

There are interdisciplinary eCRF design challenges involving technology, protocol-driven science, standardization, validation, and work-flow usability for both PDC and EDC studies. Ultimately, the final study report, which is the product of sophisticated computer programs and a statistical analysis, is only as good as the data collected in the CRF or eCRF. The whole process from defining the data to be collected, the collecting, checking, analyzing and presenting it, is resource-intensive, utilizing sophisticated technology and employing highly skilled professionals.[8] The competing/complementary demands made on the CRF or eCRF by site users, sponsors, and/or CROs must be acknowledged and addressed through balancing standards with the individual protocol requirements, considering the preference of the team members and site users, and engaging in collaboration and negotiation of the human issues involved in the process of cross-functional team review.[10] The growing importance of postmarketing data collection in large-population safety studies, the economics of drug therapy, and proteomics/genomics/pharmacogenomics presents multiple challenges including collecting, storing, integrating, querying, and analyzing growing lists of data sources, such as insurance claims, cost, large size of "omics" laboratory datasets, and patient-reported outcome data (Figure 1). It should be emphasized that study designers need to play a key role in driving and achieving core clinical database building. It is mission-critical for a sponsor to recruit a talented pool of professionals who excel in a fluid environment, pay great attention to protocol details, have developed expertise in therapeutic areas and technologies, and are capable of communicating and leveraging their working knowledge of clinical and systems engineering.[5]

Sensitive clinical operation and process re-engineering

Another challenge will be clinical process re-engineering to ensure that both PDC and EDC studies are planned, tailored, and implemented in the context of addressing clinical support, safety process improvements, and organizational needs to optimize daily clinical operations. The trend towards outsourcing continues unabated, with many organizations increasing the percentage of trials performed by CROs. When outsourcing, one must realize that the issues are not gone. Sponsor data management needs to provide guide and oversight specifically in the areas of maintaining standards and therapeutics training to ensure that CROs understand the entire clinical development spectrum and how collected data meet the efficacy and safety endpoints within the study context. It is this knowledge, collaboration, and integration that provides tangible long-term value and places a premium on having access to the right

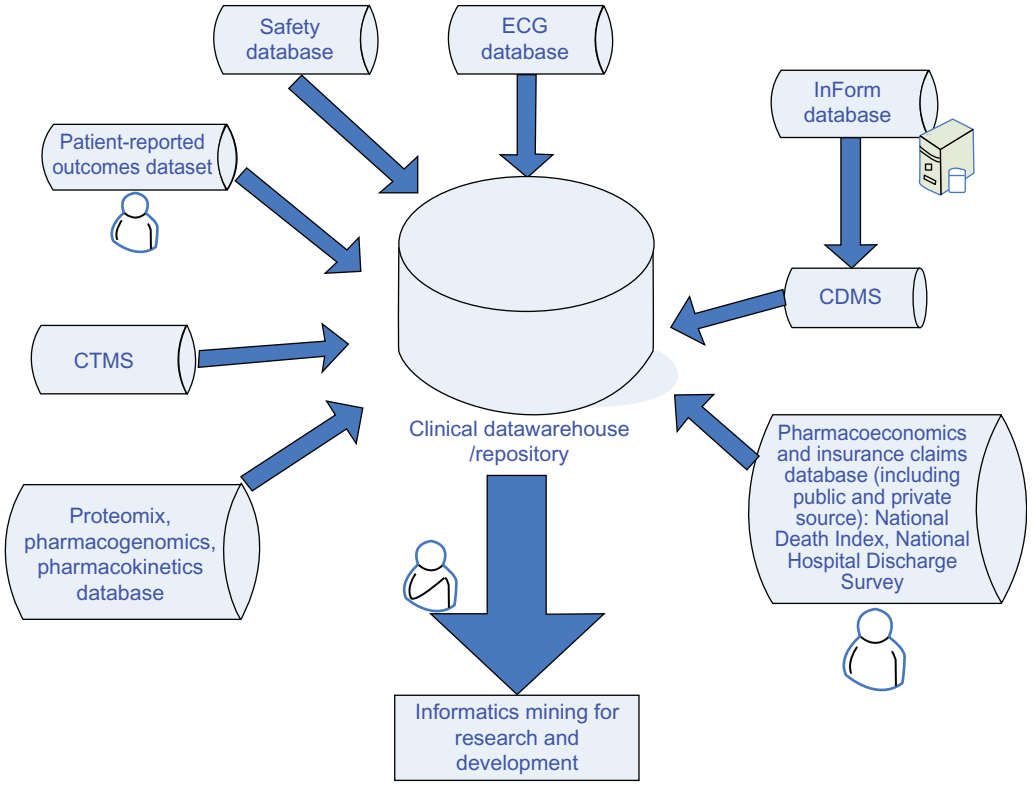

Figure 1 A sample list of potential computer data sources, database, and datasets feed data into a corporate clinical data warehouse/repository to enable informatics mining capability.
Abbreviations: ECG, electrocardiogram; CDMS, clinical data management system; CTMS, clinical trial management system.

people with the right skill set when needed. To realize the full technology-enabled benefits, the data management process needs to be re-assessed or challenged, so that redundant parts of the process can be identified and eliminated. New guidelines, business documents, or standards may be developed to support the operational needs. Table 2 summarizes key functional activities and recommended best practices under the areas of organizational alignment, operations management, and data management to enable realization of the capability of e-clinical systems in an adaptive operations framework. To address the challenges of the e-clinical environment, biopharmaceutical firms need to take flexible approaches in dealing with the legacy of paper-based procedures which exist for PDC studies only. Technology should be tapped to add process efficiencies and not to engender redundancy.[5]

Continuous technology improvement

Challenge also lies in technology improvement and flexible configurations. It is now recognized that multiple interconnected clinical systems may participate and support a clinical trial operation, indicating the absolute necessity of using contextual systems methodology when investigating and resolving any potential issues. Indeed, one must realize that current clinical systems and applications are interconnected but not interoperable, and still need reality checks on regulation and standardization. Figure 2 depicts a typical EDC data flow from sites data input through a sponsor, Clintrial, to study data tabulation model (SDTM) submission. Our experience indicates that understanding limitations and opportunities offered by an EDC vendor, configuring an EDC system to meet data-capturing needs based on a sponsor IT or data management profile, and collaborating with vendors to offer flexible configurations, are key to EDC implementation success.[10] Clearly, EDC vendors, need to take on business input, partner with industry sponsors, and offer service-oriented architecture to tackle evolving clinical research dynamics, address technology limitations, and make technology improvement. In today's technology-enabled environment, clinical data management, collaboration, and willingness to improve among multiple functional groups are key to engendering long-term clinical efficiencies and cost benefits.

Caution with edit check specifications to minimize bias

The other specific challenge for EDC concerns the "intended" higher number of autoqueries which may increase data bias.

Table 2 Proactive CDM best practices and key processes or activities in the areas of organizational alignment, operations management, and data management[2]

Organizational alignment	Operations management	Data management
Align clinical operations among both internal and external functional groups.	Implement process-driven SOPs, business documents, and training materials.	Design science-driven, site workflow-oriented, and standard-based CRF or e-CRF.
Implement decision-making, escalation processes, and communication plans.	Identify a therapeutic area subject matter expert and a CDM champion for individual operational unit.	Follow systems development life cycle methodology to design, develop, or revise a clinical database.
Implement continuous improvement and standardized risk communication plans.	Organize, review, and approve study monitoring plans in consultation with multiple other functional groups.	Develop edit check specifications,. identify, and initiate protocol specific checks.
Retain key CDM and study designer personnel with interdisciplinary expertise and skill sets.	Review, approve, and support the creation of cross-therapeutic area metrics reports to enable consistent performance measurement across studies.	Initiate, design, develop standard based metrics reports and data management reports. Initiate and implement standards across studies.
Investigate corporate IT platform and devise long-term EDC and e-clinical strategy.	Manage and provide oversight to CROs selected for portion or all of a clinical study.	Lead integration efforts in building interoperability among CDMS, CTMS, safety system, coding application.
Reorganize functional groups as needed.	Support operational issue resolutions and identify process improvement.	May be the owner of coding application, data migration, and integration.
Approve and support standards initiative.	Support standards development.	Provide training to other functional groups.

Abbreviations: CDM, clinical data management; IT, information technology; EDC, electronic data-capturing; SOP, standard operating procedure; CRO, contract research organization; CRF, case report form; eCRF, electronic case report form; CDMS, clinical data management system; CTMS, clinical trial management system.

By applying automatic data querying and controls at the point of data entry to avoid errors getting into the database, the natural data variability is drastically reduced. However, we bias the data to the desired range only, indicating potential elimination of the true data. To address this, clinical protocol design must still ensure that the study key variables will be captured by collecting the correct data. Study designers and scientists must exercise caution to ascertain that the EDC system allows site users to enter realistic data rather restrict the entry to "perfect" data only. Therefore, it is crucial that CDM and quality assurance personnel conduct thorough reviews to ensure autoquery criteria and field thresholds applied during data entry are not encouraging any data bias, and that the edit checks allow data in a truly reflective range of values so as not to overclean the data. Table 3 summarizes the list of potential issues with data entry and cleaning via EDC.

Evolving standardization and integration

Lastly, standard-based systems integration will present challenges. In the sponsor corporate environment, EDC technology and associated CDMS need to establish interoperable channels with multiple other systems, ie, IVRS/IWRS, clinical trial management system, corporate safety system, clinical coding application, and potential CRO or corporate clinical development data warehouse system or clinical data repository. Standardization of clinical protocol, common medical domains, clinical data elements, case report forms design, adverse events, and medication coding is critical to ensure quality data on study efficacy and safety assessment. Standardization is also key to ensure success of pooled data analysis among subjects in all the clinical databases used. Standardization is challenging because we do not have a standard framework yet to allow full system integration. Although the industry seems to agree that XML is the default file format for interchange and messaging, there are many implementation details to be defined and agreed to enable, for instance, a sponsor clinical study to talk directly with a hospital eHR system. It is due to this same systems interoperability challenge that current sponsor clinical studies need to collect clinical data in a separate collection instrument via eCRF or paper-based CRF although convergence is expected to continue until electronic medical or electronic health records become more pervasive within the broader health care system. At that point, the ideal solution would be to extract patient data directly from the electronic medical records as opposed to collecting the data in a separate data collection instrument or enable bidirectional channels between eHR and CDMS. Collaboration has begun in several initiatives between the CDISC,

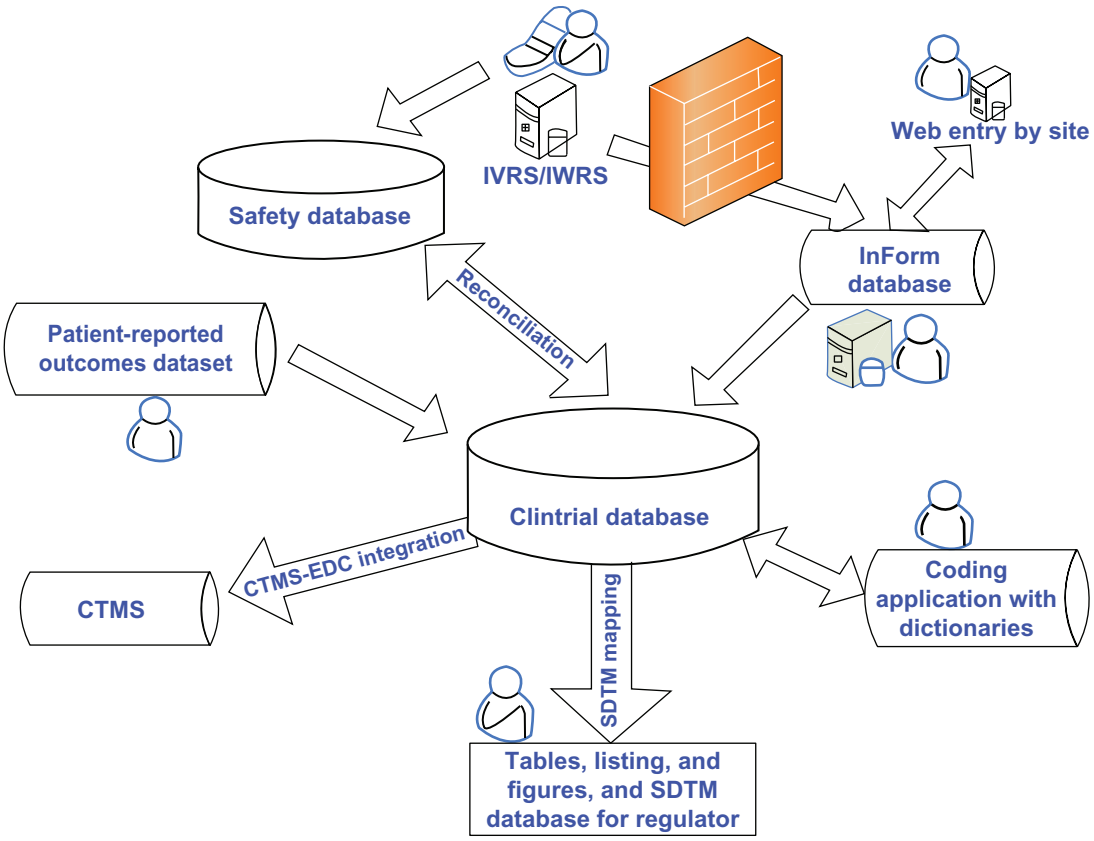

Figure 2 A typical interconnected e-clinical systems network and data workflow: site users enter data to InForm database, to sponsor Clintrial, and then transformed to SDTM format for submission.
Abbreviations: CTMS, clinical trial management system; EDC, electronic data-capturing; SDTM, study data tabulation model; IVRS/IWRS, interactive voice response system/interactive web response system.

HL7, National Cancer Institute (NCI), and FDA to encourage adoption of its global standards for clinical research, which should continue to be harmonized with health care standards, to provide a means for interoperability between health care and research systems such that clinical research can support informed health care decisions and improve patient safety.[2,11,12] Biopharmaceutical firms must adopt new processes, embrace standardizations, encourage technology innovations, retain a skilled pool of CDM talent, and undertake structured e-clinical approach and initiatives to manage the integration of technology, process and people, and be flexible to collaborate and respond in addressing issues as they occur.

Future clinical data management

The challenges to investigate clinical product candidate efficacy and safety efficiently and to adhere to regulatory

Table 3 A list of potential data entry bias associated with electronic data-capturing technology

Source of bias	Detailed description
Entry error	Site staff are not trained data entry clerks and may introduce entry errors.
Potential discrepancies between source data and eCRF entries	Data entry error might occur at sites.
Single data entry	EDC depends on single data entry. But, biopharmaceutical firms developed a double entry process as a means to improve data quality at point of entry.
Correct values that are incorrect	Site staff may fall to the EDC autoqueries to correct values that are incorrect because of their knowledge of the subject; PDC data entry staff cannot do this because they have no contact with the patient.
Overcleaning consequences	• "Impossible" values being updated • Missing values being prompted for and "constructed" • Extreme values being replaced with acceptable values • Unexpected data values being removed or modified.

Abbreviations: eCRF, electronic case report form; EDC, electronic data-capturing; PDC, paper based data collection.

requirements create the strong impression that widespread adoption of EDC technology is inevitable. Indeed, EDC and e-clinical systems have attributes attractive to the majority of biopharmaceutical firms and CROs in a competitive clinical trial industry. FDA has brought forward a critical path initiative in pushing SDTM adoption to enable electronic regulatory submissions for sponsors of human drug clinical trials. SDTM was initiated and developed by CDISC. The increasing usage of SDTM, the operational data model, analysis data model, case report tabulations data definition specification define.xml, the laboratory model, and maturing standards, such as CDASH and FDA protocols, has created an end-to-end solution for the industry to focus on moving data from the point of capture to regulatory submission, therefore boosting the adoption rate of EDC and e-clinical systems by biopharmaceutical firms. However, the apparent certainty of growing EDC adoption needs to be constantly re-examined due to considerations of a number of challenging issues.

Ongoing eHR and EDC integration

The first question is how the current standardization initiatives in reaching interoperability between differential clinical and e-health systems among several standard consortiums such as the CDISC, HL7, NCI, and FDA will play out on EDC technology.[2] The recent Initiative Electronic Health Records For Clinical Research Functional Profile has produced a functional profile to identify critical capabilities for the conduct of regulated clinical research utilizing eHR systems and additional functionalities toward facilitating ease of use for clinical research professionals.[13] Further, Roche Pharma Development and Genentech are currently conducting pilot projects[3] focused on leveraging eHR in direct support of specific drug development programs/clinical trials. These projects include concept development (mining clinical data to understand targeted patient populations better), protocol design (using current real-world clinical data to determine the impact of specific criteria on the feasibility of a protocol), and patient identification (having study sites identify potentially eligible patients directly from their eHR for proactive patient recruitment).[14] It seems promising that clinical research benefits can be realized through an eHR system. From the technical architecture perspective, will modern EDC technology system offer a multi-tier web-based application framework so that even a new clinical or health standard definition causes minimum modification? This certainly presents a challenge call to EDC vendors to partner with biopharmaceutical firms and health care technology providers to offer flexible, configurable, scalable, and interoperable EDC solutions to meet future e-clinical research needs.

Balancing technology innovation with science advancement

A second debatable question is how to balance the need for constant EDC technology, improving initiative, operational clinical support, and evolving clinical science advances. It seems reasonable that the effectiveness of the CDM function is crucial in this dynamic changing environment and hinges on science, technology, process, systems, collaboration, integration, and initiatives. Technology itself will present challenges as well as opportunities. As health care providers, health technology providers, and laboratory systems become more sophisticated and integrated, electronic data will be available from many more diverse sources and instruments. These data sources may not conform to the conventional approach of many large companies. Consequently, EDC technology and e-clinical systems have challenged traditional roles and responsibilities within clinical data management. It is increasingly realized that successful EDC implementation requires re-engineered clinical operations and culture change. Such a gear switch must obtain management support, contribution and collaboration on the part of multiple stakeholders, in which clinical science, CDM, and biostatistics play ongoing critical roles in ensuring deliverability and objectivity. Table 4 summarizes core principles for CDM to meet future challenges and what factors contribute to success in executing technology-enabled working practices and achieving quality data deliverables.

EDC technology pervasiveness with value-added cost benefit

A third unanswered question is how, exactly, the modern EDC and associated clinical systems will recruit the majority of small- to mid-sized companies, pharmacies, health care providers, and academic communities who still use labor intensive PDC tools and prefer not to change due to cost, concerns, or skepticism about EDC technology. As yet, no clear strategy has developed to assist these entities with the cost of installing, configuring, and maintaining these systems or for convincing them that they can function effectively within the new practice regimes that EDC may offer and support, with better improved return on investment compared with the PDC manual systems. Additionally, convincing top pharmaceutical companies with well established systems and processes to switch to modern sophisticated EDC systems or commit all studies to EDC

Table 4 A list of contributing factors to future challenges in clinical data management[8]

Factor	Detailed description
Embrace technology	Electronic data capture, XML, eCRF design, SQL, CDISC, CDASH, SDTM, CDMS, CTMS, pharmacovigilance, Documentum, Cognos, Oracle, SQL Server, data warehouse.
Enhance processes	The whole clinical data management process needs to be re-examined including, clinical database development, query management, and reporting. Redundant parts of the process need to be identified and eliminated.
Embed quality	The acceptable level of quality needs to be defined. Quality standards and data structures need to be promoted and applied across therapeutic areas.
Enforce regulations	Ensure compliance and to look for opportunities to improve productivity.
Mine informatics	Make sense of the data for the lifetime of the drug. Contexting information will become more important. Disciplines which generate large numbers of data points are becoming more important in medical development, such as genetics and pharmacoeconomics. The increased use of external information will also offer opportunities and challenges.
Extend communication	Electronic information exchange has become much more widespread and less complicated. This creates opportunities, both in improved process and in facilitating better two-way communication.
Expand resource	Resource the data management function to build a diversified pool of talent with differing skill sets.
Empower data managers	Different skill sets are emerging. The challenge is to build the skills, recruit and retain good people.
Evolve culture	Nurture a culture which attracts the right kind of employees to fulfill this important role.
Extend to emerging markets	Work in such regions to build a relevant clinical information source and to put effective data management structures into place.

Abbreviations: XML, extensible markup language; eCRF, electronic case report form; SQL, structured query language; CDISC, clinical data interchange standards consortium; CDASH, clinical data acquisition standards harmonization; SDTM, standard data tabulation model; CDMS, clinical data management; CTMS, clinical trial management system.

can be both challenging and exciting. One needs to possess at least the following assets to succeed: ability to demonstrate enhanced system functionality and configurability, an understanding of business requirements, a commitment to customer service, ability to assist with data migration and system knowledge transfer, ability to offer consultation in preparation of new standard operating procedures or modification of existing ones, and ability to demonstrate cost-saving advantages in the long-term. The most difficult item seems to be aligning or adjusting existing processes to fit into the new system.[2]

Science-driven standard-based clinical development data warehouse

The biggest uncertainty concerning EDC technology and e-clinical systems is how much data warehousing or integration effort is required for a sponsor to take advantage of the vast variety and huge amount of data available, including (but not limited to) clinical data collected via CRF or eCRF, data captured through e-diaries, laboratory data generated via 2D or 3D imaging diagnostics, produced via central laboratory instrumentation, safety data stored in corporate safety system, patient data captured via eHR, "omics" data accumulated in translational research spectrum and how mining such data may break through the barriers that constrain productivity to bring new insights into the study of disease and human populations.[15] Such challenging development may be an appropriate option for some biopharmaceutical firms only. Undertaking such enterprise level initiative requires top management vision, accountable resource or consulting commitment, a long-term clinical development strategy, and close partnership among all therapeutic units. Effectively translating this knowledge into clinical intelligence and improved patient care and efficient utilization of such vast informatics data are holding potentials to advance the conduct of science and design new clinical programs for future medicine. Such novel strategies based on multiple sources of data attributes may open up new opportunities, transform how clinical medicine is practiced, and offer earlier interventional measures in the treatment process to stop diseases before they occur. The framework for this data-driven personalized vision is centered on the model of predictive, personalized, preemptive, and participatory medicine. Practicing medicine in this way will help us move more quickly to understand the fundamental causes of diseases at their earliest molecular stages so that we can reliably predict how, when, and in whom a disease will develop due to individual genetic compositions and difference in response to environmental changes/stresses. In order to realize this individualized approach and incorporate informatics into a sponsor data warehouse, the rigor to improve and innovate will be primary, the standardization and integration secondary, and patience and collaboration

critical. Creating a standards-based and interoperable clinical development data repository/data warehouse in which corporate management, clinical science and safety staff can perform data mining and quality improvement in identifying process optimization, setting clinical product candidate priority, detecting safety signal, and reducing cost to accomplish corporate financial and professional goals will be paramount to widespread adoption of modern EDC technology and e-clinical systems and to assessing their transformative potential.[2,8]

Conclusion

The competitive pressure in today's marketplace is forcing the biopharmaceutical industry to seek better ways of reducing drug development times and increasing productivity. The market acceptance of EDC technology has fueled new demands for improvement, configurability, and intelligent features.[5] The need to improve clinical efficiencies and accelerate study times continues to grow, driving industry sponsors to seek an e-clinical environment that provides and promotes flexible eCRF trial design, build, and speedy deployment, robust data management, real-time data visibility, reporting and analysis, and global trial management and study scalability.[10] Shortening the clinical trial lifecycle by collecting quality data more quickly and accelerating the availability of data are solutions to a critical path bottleneck that the industry has been working on for many years.[16] Adopting EDC technology and e-clinical systems in the clinical trial process offers a solution with some claimed success stories. This has led to the growth of a new industry of clinical software vendors, offering a host of systems from EDC to IVRS, ePROs to CTMS, central coding application to safety signal detection, and clinical data warehouse initiatives to race towards e-clinical realization. The availability of near-real time data through the use of EDC has opened the door to the development of an integrated e-clinical environment. Yet, PDC-based clinical studies represent a fair percentage of studies in many organizations.[17] Where EDC is being used at scale, operational benefits are being realized. The near-real time data, increasing standardization among multiple stakeholders, and integrated clinical environments have produced a paradigm shift in the clinical development model from research hypothesis, patient experience, through to analysis and SDTM submission. EDC technology and e-clinical systems have the potential to meet the challenges of providing powerful support to identify and discover the increasing range and potency of medicines. However, there are issues, concerns, and challenges in implementing and configuring modern EDC solutions. Clinical research professionals need to anticipate proactively, embrace attentively, and prepare for the further diversified challenges from both systems and business engineering perspectives in the world of Internet medicine.[18]

Disclosure

The authors report no conflicts of interest in this work. Moreover, opinions or views expressed through this article represent individual perspectives only.

References

1. Pavlovic I, Kern T, Miklavcic D. Comparison of paper-based and electronic data collection process in clinical trials: Costs simulation study. *Contemp Clin Trials.* 2009;30:300–316.
2. Lu ZW. Information technology in pharmacovigilance: Benefits, challenges, and future directions from industry perspectives. *Drug, Healthcare and Patient Safety.* 2009;1:35–45.
3. EI Emam K, Jonker E, Sampson M, et al. The use of electronic data capture tools in clinical trials: Web-survey of 259 Canadian trials. *J Med Internet Res.* 2009;11:e8.
4. US Food and Drug Administration. Guidance for Industry: Computerized Systems used in Clinical Investigations, May 2007. Available from: http://www.fda.gov/OHRMS/DOCKETS/98fr/04d-0440-gdl0002.PDF. Accessed March 23, 2010.
5. Lu ZW. Electronic data-capturing technology for clinical trials: Experience with a global postmarketing study. *IEEE Eng Med Biol Mag*. 2010;29:95–102.
6. Mosley M. DAMA-DMBOK Guide (data management body of knowledge) Introduction and Project Status, November 2007. Available from: http://www.dama.org/files/public/DI_DAMA_DMBOK_Guide_Presentation_2007.pdf. Accessed March 23, 2010.
7. DAMA-DMBOK Editorial Board. *The DAMA Guide to the Data Management Body of Knowledge.* Bradley Beach, NJ: Technics Publications; 2009.
8. Rondel RK, Varley SA, Webb CF. *Clinical Data Management.* 2nd ed. West Sussex, UK: John Wiley and Sons Ltd; 2000.
9. Ishigaki D. Effective Management through Measurement, May 2004. Available from: http://www.ibm.com/developerworks/rational/library/4786.html. Accessed March 23, 2010.
10. Lu ZW. Technical challenges in designing post-marketing eCRFs to address clinical safety and pharmacovigilance needs. *Contemp Clin Trials.* 2010;31:108–18.
11. Clinical Data Interchange Standards Consortium (CDISC). CDASH Core and Domain Teams. *Clinical Data Acquisition Standards Harmonization (CDASH)*. Austin, TX: Clinical Data Interchange Standards Consortium; October 2008.
12. Buetow K. Building a 21st Century Biomedical System: The Cancer Biomedical Informatics Grid (caBIG®). Annual meeting of DIA data conference. Philadelphia, PA, March 9–11, 2009.
13. King L, Bishop S, Perkins R, et al. EHRCR User Requirements Document, January 2010. Available from: http://www.ehrcr.org/Docs/EHRCR%20User%20Req%20Rel%202%20Jan%202010.pdf. Accessed March 24, 2010.
14. Myles J. Experiences leveraging Electronic Health Records (EHR) in direct support of drug development. CHI's 2nd Annual Electronic Data in Clinical Trials: Collecting and leveraging data to optimize clinical trials. Philadelphia, PA, March 8–9, 2010.
15. Young DW, Tao CY, Feng Y. Data analysis pipeline for high content screening in drug discovery. In: Rao AR, Cecchi GA, editors. *High-Throughput Image Reconstruction and Analysis*. Norwood, MA: Artech House; 2009.

16. Forster E. The changing R and D model – What does e-clinical development need to support? 2006. Available from: http://www-935.ibm.com/services/uk/bcs/pdf/eclinical_equation.pdf. Accessed March 24, 2010.
17. Connelly DM. The Challenges of Data Capture and Data Management – a CRO Perspective, 2006. Available from: http://www-935.ibm.com/services/uk/bcs/pdf/eclinical_equation.pdf. Accessed March 24, 2010.
18. Lehoux P, Daudelina G, Demers-Payettea O, et al. Fostering deliberations about health innovation: What do we want to know from publics? *Soc Sci Med.* 2009;68:2002–2009.